

The effect of DNA extraction methodology on gut microbiota research applications

Gerasimidis, Konstantinos; Bertz, Martin; Quince, Christopher; Brunner, Katja; Bruce, Alanna; Combet, Emilie; Calus, Szymon; Loman, Nick; Ijaz, Umer Zeeshan

DOI:

[10.1186/s13104-016-2171-7](https://doi.org/10.1186/s13104-016-2171-7)

License:

Creative Commons: Attribution (CC BY)

Document Version

Publisher's PDF, also known as Version of record

Citation for published version (Harvard):

Gerasimidis, K, Bertz, M, Quince, C, Brunner, K, Bruce, A, Combet, E, Calus, S, Loman, N & Ijaz, UZ 2016, 'The effect of DNA extraction methodology on gut microbiota research applications', *BMC Research Notes*, vol. 9, 365. <https://doi.org/10.1186/s13104-016-2171-7>

[Link to publication on Research at Birmingham portal](#)

General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

RESEARCH ARTICLE

Open Access



The effect of DNA extraction methodology on gut microbiota research applications

Konstantinos Gerasimidis¹, Martin Bertz¹, Christopher Quince², Katja Brunner¹, Alanna Bruce¹, Emilie Combet¹, Szymon Calus³, Nick Loman⁴ and Umer Zeeshan Ijaz^{3*}

Abstract

Background: The effect that traditional and modern DNA extraction methods have on applications to study the role of gut microbiota in health and disease is a topic of current interest. Genomic DNA was extracted from three faecal samples and one probiotic capsule using three popular methods; chaotropic (CHAO) method, phenol/chloroform (PHEC) extraction, proprietary kit (QIAG). The performance of each of these methods on DNA yield and quality, microbiota composition using quantitative PCR, deep sequencing of the 16S rRNA gene, and sequencing analysis pipeline was evaluated.

Results: The CHAO yielded the highest and the QIAG kit the lowest amount of double-stranded DNA, but the purity of isolated nucleic acids was better for the latter method. The CHAO method yielded a higher concentration of bacterial taxa per mass (g) of faeces. Sequencing coverage was higher in CHAO method but a higher proportion of the initial sequencing reads were retained for assignments to operational taxonomic unit (OTU) in the QIAG kit compared to the other methods. The QIAG kit appeared to have longer trimmed reads and shorter regions of worse quality than the other two methods. A distinct separation of α -diversity indices between different DNA extraction methods was not observed. When compositional dissimilarities between samples were explored, a strong separation was observed according to sample type. The effect of the extraction method was either marginal (Bray–Curtis distance) or none (unweighted Unifrac distance). Taxon membership and abundance in each sample was independent of the DNA extraction method used.

Conclusions: We have benchmarked several DNA extraction methods commonly used in gut microbiota research and their differences depended on the downstream applications intended for use. Caution should be paid when the intention is to pool and analyse samples or data from studies which have used different DNA extraction methods.

Keywords: Metagenomics, DNA extraction, Benchmarking, Diversity, PCR, 16S rRNA gene

Background

The introduction of molecular biology techniques and deep sequencing in microbiology has revolutionised our interest and advanced our understanding on the role of gut microbiota in health and disease. Isolation and purification of bacterial genomic DNA from gut mucosal and luminal contents is a crucial initial step to

ensure a high yield and quality of isolated nucleic acids, and unbiased representation of microbial communities. Over the past decade, several proprietary DNA extraction kits have been developed and became commercially available with the intention to replace the more laborious, time consuming original approaches [1, 2]. There is good evidence showing that different DNA extraction kits will generate different results in terms of: amount and quality of extracted DNA, inhibitors of PCR reactions, and influences on bacterial community composition [1–5]. The effects that various DNA

*Correspondence: Umer.Ijaz@glasgow.ac.uk

³ School of Engineering, University of Glasgow, Oakfield Avenue, Glasgow G12 8LT, UK

Full list of author information is available at the end of the article

extraction methods may have on traditional downstream methods (e.g. quantitative PCR) compared to modern next generation sequencing approaches have not been extensively explored [1, 2]. Hereby, we have performed a benchmarking study and explored the effect of three popular faecal DNA extraction methods on yield and quality of isolated DNA, as well as microbiota composition determined by a typical sequencing analysis pipeline, using traditional molecular microbiology techniques and high-throughput next generation sequencing.

Results

Effect on yield and purity of isolated DNA

Beginning with the same faecal material mass, the chaotropic method (CHAO) produced the highest and the Qiagen kit (QIAG) the lowest yield of double stranded DNA (Fig. 1a). However, for the absorbance ratio at 260/280 nm, a metric of isolated nucleic acids purity, the QIAG kit performed best and the CHAO method worst (Fig. 1a). Extraction of a proprietary probiotic capsule with the phenol/chloroform (PHEC) method gave the highest yield of double-stranded DNA but this was comparatively of very poor quality.

Effect on quantitative estimates of major bacterial taxa

Independently of faecal sample or bacterial taxon explored, the same mass of faecal sample gave a higher concentration (per g of original faeces) of 16S rRNA amplicon copies with the CHAO method in qPCR analysis (Fig. 1b). For the probiotic capsule, the PHEC method produced the highest concentration of amplicon copies for all taxa identified.

Effect on microbial community determined through 16S rRNA amplicon sequencing

The mean (SD) number of reads per method was higher for the CHAO and PHEC than the QIAG method (Additional file 1: Figure S1, Additional file 2: Table S1). In terms of sequencing yield, more reads from the QIAG kit were obtained for OTU construction (Fig. 2a). In accordance to our recent publication, the reverse reads for the Illumina platform were of worse quality than the forward ones [6] and hence the difference in the number of reads after quality trimming. Trimming improved the read quality to an extent and sequencing reads originated from the PHEC method had better quality than the other two approaches (Fig. 2b). In terms of read length, the QIAG kit appeared to have longer trimmed reads and shorter regions of worse quality, than the other two methods. The PHEC method produced the longest overlapped reads (Fig. 2c).

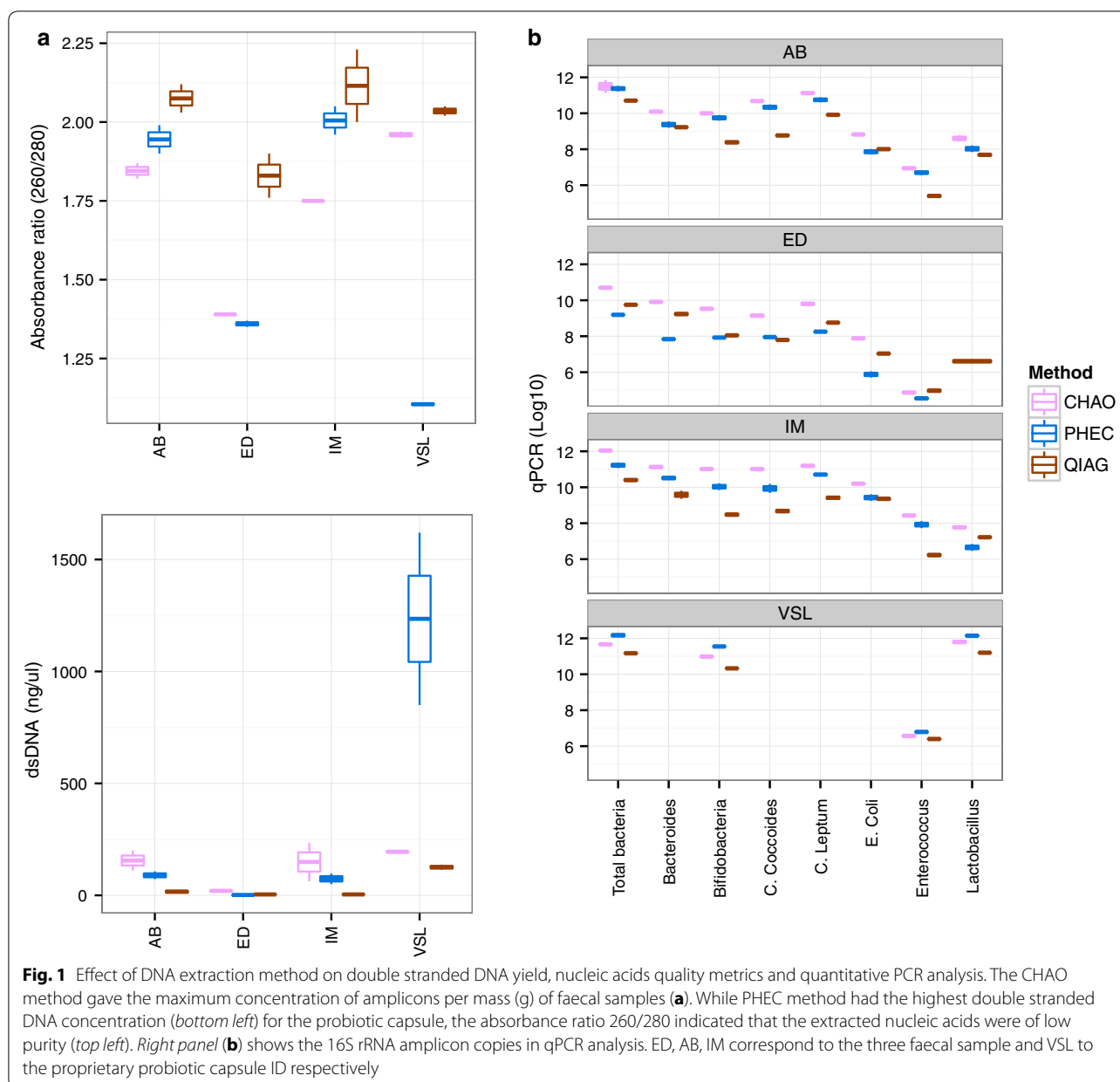
Effect on 16S rRNA amplicon sequencing based community composition

For a range of α -diversity measures, a distinct pattern was not observed for different extraction methods (Fig. 3). Using different β -diversity measures for between sample similarities, on non-metric multidimensional scaling plots, a clear separation was observed between the sample types but not according to the extraction methods (Fig. 4a, b). This was the same when considering distances between samples calculated on OTU abundance counts (Bray–Curtis distance) or on their phylogenetic relatedness (unweighted Unifrac) (Fig. 4a, b). These visual cues were then confirmed using PERMANOVA (using vegan's adonis package in R) analysis with sample type accounting for 51 % ($p = 0.001$) and extraction method explaining 11 % ($p = 0.023$) of the variation in community structure using the Bray–Curtis dissimilarity index. However, when unweighted Unifrac distance was used, the majority of the variation ($R^2 = 0.75$; $p = 0.001$) was attributed to sample type, with extraction method becoming non-significant ($R^2 = 0.049$; $p = 0.114$). Using a recently proposed diversity estimator (BAT package in R), the technical replicates (a measure of method reproducibility) were closer to each other for the QIAG and CHAO methods than the PHEC method (Fig. 4c, d).

Even though each sample had distinct community profiles, taxon membership in each of them was independent of the extraction method used (Fig. 5). Similarly, no obvious differences were observed between the three methods at taxon abundance (Fig. 6).

Discussion

In this study we have observed that different DNA extraction methods for gut microbiota research applications lead to different results in downstream data analysis. No DNA extraction method was superior or inferior to another and the implications of their use will depend on the type of applications considered (Table 1). When the objective is to estimate the absolute concentration of the bacterial taxa in applications such as qPCR, the usage of the CHAO method will give higher values per mass of faecal material compared with the QIAG kit; a result which was merely explained by the higher double stranded DNA yield of the former method. The loss of nucleic acids in the QIAG kit is difficult to ascertain but this is likely to be associated with loss of the nucleic acids within the spin column during the purification process step; albeit the latter step might improve the quality of resulting nucleic acids. These results preclude comparison of qPCR data produced from samples extracted with different methods. The higher yield of the CHAO may also be of relevance in microbiota research with specimens of very low bacterial biomass, as in profiling of the

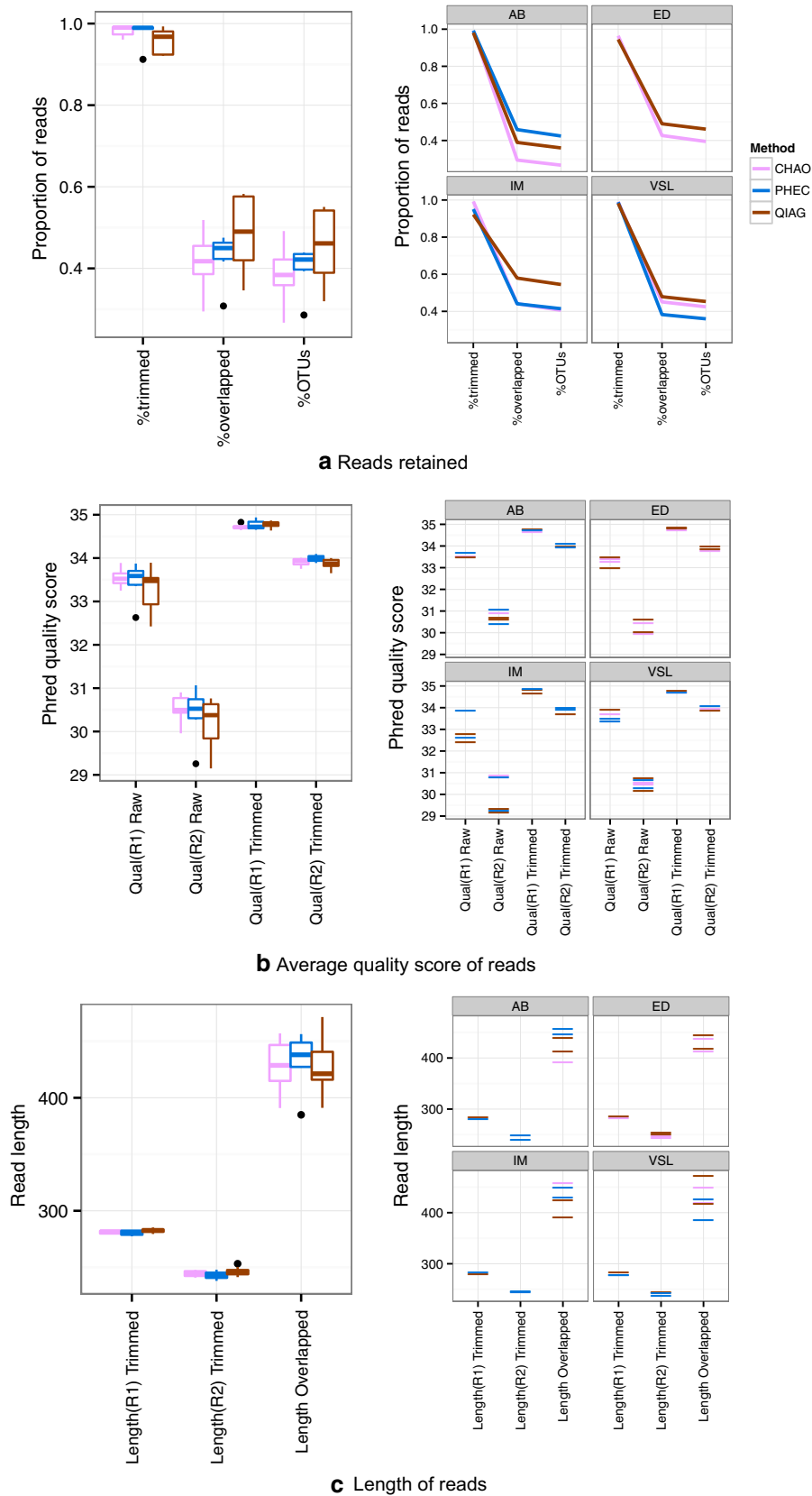


mucosal microbiota. However, when the research objective is to study the community structure and estimate the relative abundance of taxa, using next generation sequencing of the 16S rRNA amplicon gene, no substantial differences are expected to be observed between the various DNA extraction approaches tested here.

From a bioinformatics analysis perspective, an important determinant of a suitable DNA method is the high quality sequencing yield. A method which produces a high number of reads per sample, and with a high percentage of reads mapping to OTUs following bioinformatics analysis is desirable. Such a method will be more

(See figure on next page.)

Fig. 2 Effect of DNA method on each step of typical bioinformatics analysis pipeline. The *left* and *right* columns show organisation by per method and per sample basis, respectively. In terms of the total useful reads that are mapped to OTU, QIAG kit performed better than the other two methods as roughly 50 % of the reads were mapped to OTUs during the construction stage (see Additional file 2 : Table S1 for actual read numbers and Additional file 1 : Figure S1 summary statistics). In terms of quality trimming, QIAG lost more reads than the other two methods. ED, AB, IM correspond to the three faecal sample and VSL to the proprietary probiotic capsule ID respectively



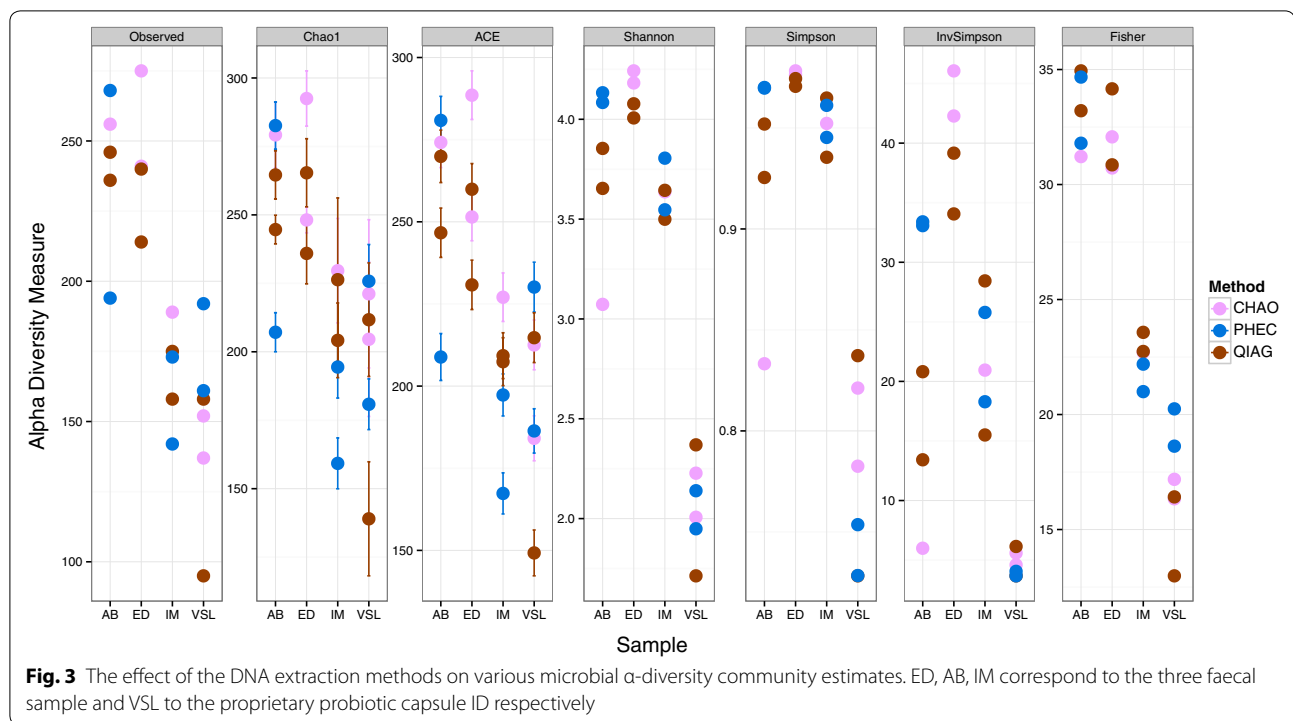


Fig. 3 The effect of the DNA extraction methods on various microbial α -diversity community estimates. ED, AB, IM correspond to the three faecal sample and VSL to the proprietary probiotic capsule ID respectively

useful for statistical analysis and will be more cost-effective by decreasing the number of repeated sequencing analysis of some samples (often a library size cut off is applied to filter out samples with low abundance count). In this context, the QIAG was superior over the two other approaches.

The main limitation of this study is the small number of samples which precluded formal statistical analysis for some outcome measures. However, even with this small number of samples the majority of the results were consistent and of the same direction. Moreover, the results of this study should be interpreted with relevance to faecal specimens and the performance of these methods in other matrices (e.g. soil or plants) needs to be explored.

As research in the area of gut microbiology is moving from small scale to large multicentre international studies, a choice of practical, quick and cost effective methods will be preferred. In this instance, the CHAO method takes up to 2 days to process 15 samples at an average cost of less than £1 per sample in consumables, whereas the PHEC extraction can be completed in less than 6 h, at the same average cost as the CHAO method, and the QIAG kit will only take 3 h at an approximate cost of £4 per sample for the same number of samples and can be incorporated within automated DNA extractors.

Conclusions

In conclusion, we have shown that there is no superior DNA method fitting all downstream approaches and the method of choice depends on the intended type of analysis, practicality and cost (Table 1). Nonetheless, we advocate towards the importance of using the same DNA extraction method when comparing group differences in a study as well as caution should be paid when the intention is to analyse biobanked samples or pool data from studies which have used different DNA extraction methods.

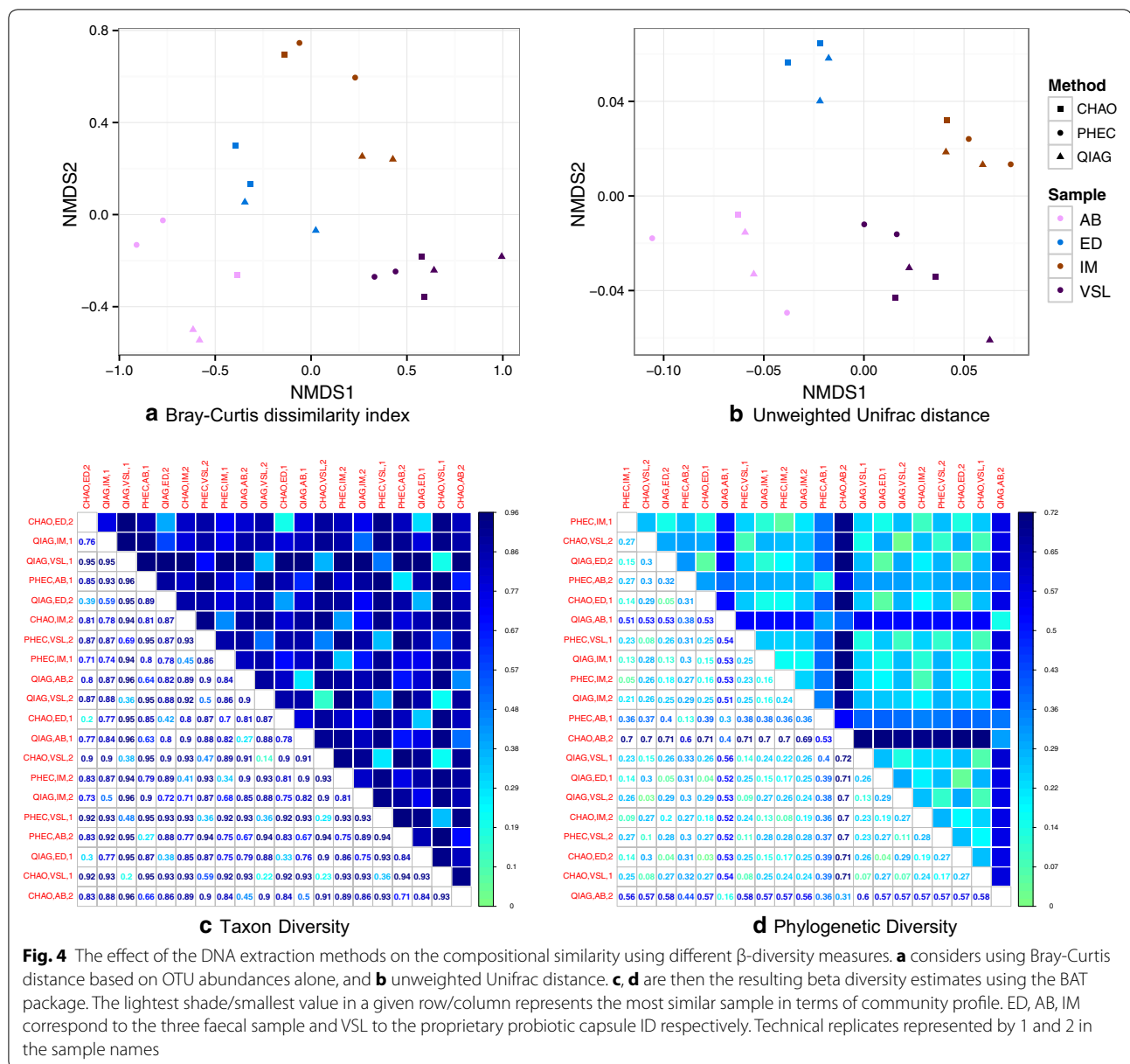
Methods

Sample collection

Faecal samples were obtained from a healthy adult, a healthy child, and a child with Crohn's disease who participated in ongoing research [7]. A proprietary probiotic preparation containing 8 different strains of lactic-acid bacteria, VSL#3[®] (sigma-tau Ethifarma b.v. NL), was analysed too.

DNA extraction methods

Extraction of 200 mg stool was carried out in duplicate for each method (technical replicates) as described below. Three different DNA extraction protocols were used: (a) a modified commercial kit (QIAamp[®] DNA stool mini kit, Qiagen), (b) a phenol–chloroform extraction method and



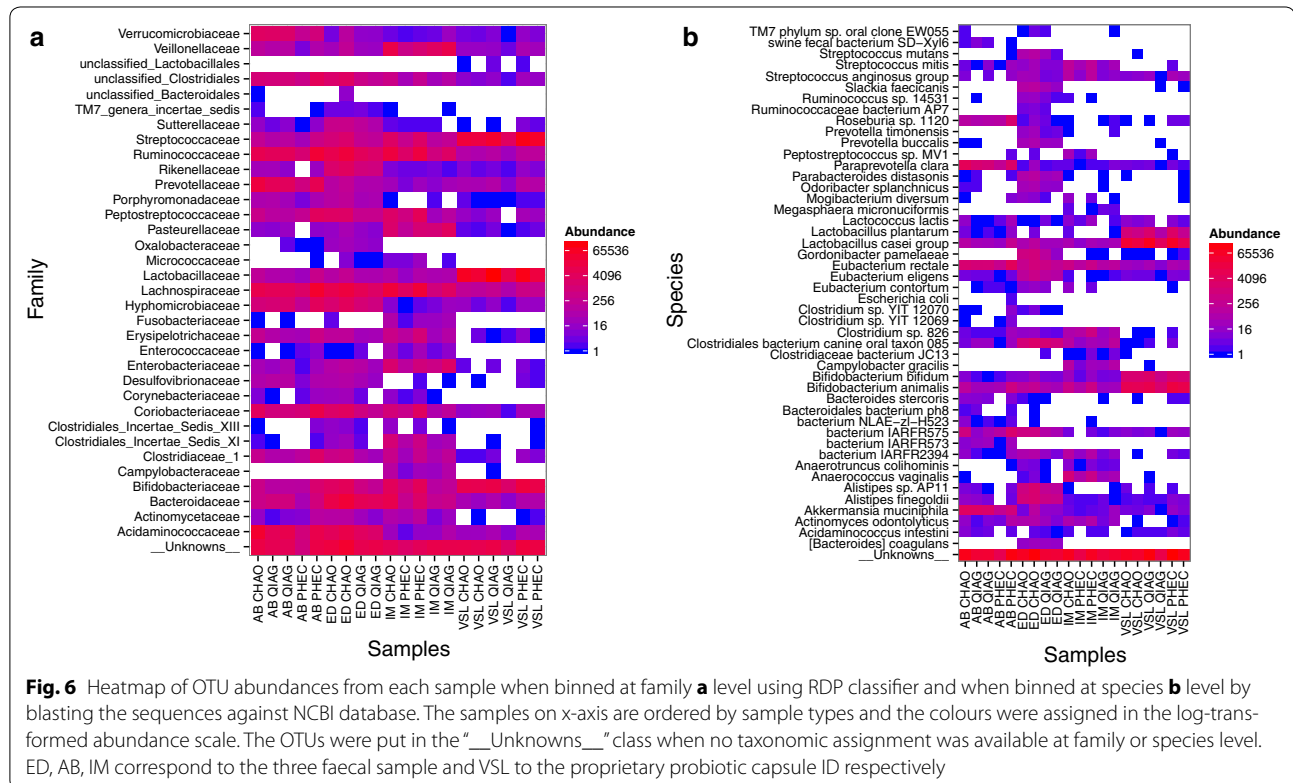
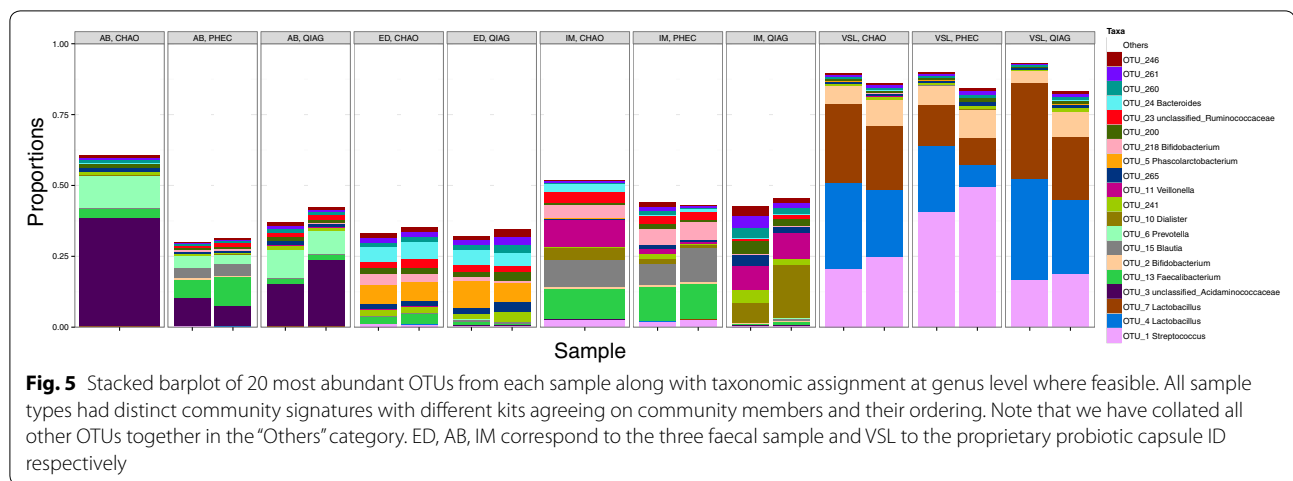
c) a method using a combination of chemical, enzymatic and physical steps.

QIAamp® DNA stool mini kit

The kit was applied according to the manufacturer's instructions with modifications. These included destruction of bacterial cells with bead beating (Tungsten Carbide Beads 3 mm Cat No. 699997, QIAGEN) for 3 min at 4.5 m/s with a FastPrep-24 (MP biomedical).

Phenol–chloroform extraction

The phenol–chloroform extraction used is based on the protocol by Reichardt et al. [8]. This involved mechanical lysis destruction of bacterial cells with zirconium beads (0.1 mm, Biospec products) in sterile PBS, saturated acid phenol and separation of nucleic acids with chloroform/isoamylalcohol (24:1), centrifugation and precipitation of nucleic acids with isopropanol occurred in the presence of 3 M sodium chloride.



Chaotropic method

A modified version of the protocol by Godon et al. [9] as described previously [10]. Briefly faeces were suspended in a buffer containing a salt solution and incubated for 1 h at 70 °C. Sterile silica beads (0.1 mm, Biospec products) were used for bacterial cell lysis in a FastPrep-24 bead beater (MP biomedical). Then 15 mg Polyvinylpyrrolidone was added and the suspension was centrifuged with 15,000×g at 4 °C for 3 min. The supernatant was

recovered, the pellet was washed with 450 µL TENP buffer, centrifuged again and washed two more times. The pooled supernatants were centrifuged with 20,000×g at 4 °C for 10 min. Nucleic acids were precipitated with isopropanol. Following 10 min incubation at room temperature the mixture was centrifuged with 15,000×g at 4 °C for 5 min and the supernatant was discarded. The pellet was resuspended in 225 µL phosphate buffer 0.1 M (pH 8) and 25 µL potassium acetate 5 M and left at 4 °C

Table 1 Summary of findings reported in this study

| Methods | Pros | Cons |
|---------------------------------|---|--|
| Chaotropic (CHAO) | Highest DNA yield | Most fractional loss of reads after different steps of bioinformatics pipeline |
| | High coverage in sequencing | Takes longest to perform |
| | Lowest loss of read trimming | |
| | Cheap | |
| Phenol Chloroform (PHEC) | Cheap | In the middle in terms of performance compared with the other methods |
| | Highest dsDNA in probiotic sample | Use toxic reagents |
| | Quicker than CHAO | |
| QIAamp DNA stool minikit (QIAG) | Highest read yield post OTUs assignment | Lowest DNA yield |
| | Quick | Not cheap |
| | Highest quality DNA | Lowest coverage |
| | Can be automated | |

overnight. 5 µL RNase (10 mg/mL) was added and incubated at 37 °C for 45 min. DNA was precipitated using 50 µL 3 M sodium acetate and 1 mL ice cold 100 % ethanol. After incubation at −20 °C for 1 h the DNA pellet was washed three times with 70 % ethanol, dried and stored at −20 °C in TE buffer.

Yield and purity of isolated nucleic acids

Bacterial genomic double stranded DNA yield was measured with the Qubit® fluorometer 2.0 using the high sensitivity assay kit (ThermoFisher, Q32851) and the purity of nucleic acids was assessed with at the absorbance ratio 260:280 (NanoDrop® ND-1000).

Quantitative real time PCR

The concentration of 16s rRNA gene copies of major dominant and subdominant bacterial taxon groups (*Clostridium leptum*, *Clostridium coccooides*, *Bifidobacterium genus*, *Lactobacillus*, *Escherichia coli*, *Enterococcus*) were measured in triplicate using quantitative real-time PCR analysis on a 7500 Real-Time PCR System (Applied Biosystems) using TaqMan Gene Expression and the same primers and probes as described previously [10]. The concentration of 16S rRNA gene copy number for each sample was expressed per gram of dry faecal material taking into account any dilution factor in the concentration of template DNA in qPCR reaction. Non template controls were included in each run.

16S rRNA gene sequencing

16S rRNA gene sequencing was performed on the MiSeq (Illumina) platform using 2 × 250 bp paired-end reads. The V4 region was amplified using fusion Golay adaptors barcoded on the reverse strand as described previously [7]. The forward 16S rRNA primer sequence 515f

(GTGCCAGCMGCCGCGGTAA) was used. The reverse primers, barcodes and adaptors were identical to those described previously [11]. Amplicons were purified with AMPure XP DNA purification beads (Beckman Coulter, Danvers, MA, USA) according to the manufacturer's instructions, and eluted in 25 µl of proprietary elution buffer (Qiagen, 19086, UK). Amplicon concentration was quantified with use of KAPA SYBR® FAST qPCR Kit (Kapa biosystems, KK4824, UK), diluted to 40 pM and spiked with 40 pM of genomic DNA to avoid base-calling issues due to low base diversity. A negative extraction control was included for each method.

Bioinformatics

The paired end reads were trimmed and filtered using Sickle v1.200 [12] by applying a sliding window approach and trimming regions where the average base quality drops below 20. After this, we applied a 10 bp minimum length threshold to discard all shorter reads. We then used pandaseq v(2.4) [13] with a minimum overlap of 50 bp to assemble the forward and reverse reads into a single sequence spanning the entire V4 region. After obtaining the consensus sequences from each sample, we used the UPARSE (v7.0.1001) pipeline (<https://bitbucket.org/umerijaz/amplimock/src>) for OTU construction. In brief, we pooled reads from different samples together and added barcodes to keep an account of the samples these reads originated from. We then dereplicated the reads and sorted them by decreasing abundance and discarded singletons. In the next step, the reads were clustered based on 97 % similarity discarding reads that were shorter than 32 bp. Even though the cluster_otu command in usearch removes reads that have chimeric models built from more abundant reads, a few chimeras may be missed, especially if they have parents that are absent

from the reads or are present with very low abundance. Therefore, in the next step, we used a reference-based chimera filtering step using a gold database (http://drive5.com/uchime/uchime_download.html) that is derived from the ChimeraSlayer reference database in the Broad Microbiome Utilities (<http://microbiomeutil.sourceforge.net/>). The original barcoded reads were matched against clean OTUs with 97 % similarity (a proxy for species level separation) to generate a total of 335 OTUs comprising all samples. The representative OTUs were then taxonomically classified against the RDP database using the standalone RDPclassifier v2.6 [14] with the default—min-Words option of 5. For species level assignment, we have used NCBI Taxonomy and TAXAassign (<https://github.com/umerijaz/TAXAassign>). To find the phylogenetic distances between OTUs, we first multisequence aligned the OTUs against each other using mafft v7.040 [15] and then used FastTree v2.1.7 [16] on these alignments to generate an approximately-maximum-likelihood phylogenetic tree.

Statistical analysis

Statistical analyses were performed in R using the tables and data generated as above as well as the metadata associated with the study. For community analysis (including α and β diversity analyses) we have used the vegan [17] package in R. To obtain unweighted Unifrac distances (that account for phylogenetic relatedness and are calculated using the branch lengths from the phylogenetic tree of the OTUs observed in the samples, without considering their abundances), we have used the phyloseq [18] package. Non-metric distance scaling plot (NMDS) was applied using Vegan's metaMDS() function to visualise natural clustering in the dataset. Additionally, we have used ape [19], phangorn [20] and BAT [21] packages together to calculate the recently proposed β -diversity estimators that consider phylogeny too. The BAT package proposes three ways of estimating phylogenetic diversity (PD), building on estimators originally developed for Taxon Diversity (TD): correcting PD values based on the completeness of TD; fitting asymptotic functions to accumulation curves of PD; and adapting nonparametric estimators to PD data. The only requirement is for the phylogenetic tree to be an ultrametric tree for which we used chronos() from R's ape package to convert our OTU tree to an ultrametric tree (after rooting the tree by applying midpoint() rooting function from the R's phangorn package). In Fig. 4c, d, the resulting beta diversity estimates from BAT packaged are plotted using R's corplot to represent the quantitative estimates and then ordered using R's hclust(). The general scripts as well as tutorials for the above analyses are available at <http://userweb.eng.gla.ac.uk/umer.ijaz#bioinformatics>.

Additional files

Additional file 1: Figure S1. Summary statistics of read qualities and read lengths after each step of the pipeline and before OTUs construction.

Additional file 2: Table S1. Sample raw data of read qualities and read lengths after each step of the pipeline and before OTUs construction.

Authors' contributions

KG, UZI, wrote the manuscript, KG, MB, SC, NL processed the samples, UZI, KG analyzed the data, NL, CQ, EC, critically discussed the findings and gave feedback on the manuscript. KG co-ordinated the project, provided funding. All authors read and approved the final manuscript.

Author details

¹ Human Nutrition, School of Medicine, College of Medical, Veterinary and Life Sciences, Glasgow Royal Infirmary, University of Glasgow, Glasgow, UK. ² Warwick Medical School, University of Warwick, Warwick, UK. ³ School of Engineering, University of Glasgow, Oakfield Avenue, Glasgow G12 8LT, UK. ⁴ Institute of Microbiology and Infection, University of Birmingham, Birmingham B15 2TT, UK.

Acknowledgements

C. Quince is funded by an MRC fellowship as part of the CLIMB consortium Grant Ref: MR/L015080/1. U.Z. Ijaz is funded by a NERC fellowship NE/L011956/1. We would like to thank the anonymous reviewer 3 for his suggestion to summarise findings of the study within a Table.

Competing interests

The authors declare that they have no competing interests.

Availability of supporting data

The raw sequence files supporting the results of this article are available in the EBI repository, (deposited upon acceptance). The data are available on the European Nucleotide Archive under the study accession number: PRJEB14875 (<http://www.ebi.ac.uk/ena/data/view/PRJEB14875>).

Ethics approval and consent to participate

This study has received full ethical permission by the Yorkhill Research Ethics Committee (Reference Number: 05/S0707/66) and each participant offered written informed consent according to NHS Good Clinical Practice for Research.

Received: 12 May 2016 Accepted: 19 July 2016

Published online: 26 July 2016

References

- Kennedy NA, Walker AW, Berry SH, et al. The impact of different DNA extraction kits and laboratories upon the assessment of human gut microbiota composition by 16S rRNA gene sequencing. *PLoS ONE*. 2014;9:e88982.
- Salonen A, Nikkila J, Jalanka-Tuovinen J, et al. Comparative analysis of fecal DNA extraction methods with phylogenetic microarray: effective recovery of bacterial and archaeal DNA using mechanical cell lysis. *J Microbiol Methods*. 2010;81:127–34.
- de Boer R, Peters R, Gierveld S, et al. Improved detection of microbial DNA after bead-beating before DNA isolation. *J Microbiol Methods*. 2010;80:209–11.
- McOrist AL, Jackson M, Bird AR. A comparison of five methods for extraction of bacterial DNA from human faecal samples. *J Microbiol Methods*. 2002;50:131–9.
- Nechvatal JM, Ram JL, Basson MD, et al. Fecal collection, ambient preservation, and DNA extraction for PCR amplification of bacterial and human markers from human feces. *J Microbiol Methods*. 2008;72:124–32.

6. D'Amore R, Ijaz UZ, Schirmer M, et al. A comprehensive benchmarking study of protocols and sequencing platforms for 16S rRNA community profiling. *BMC Genom.* 2016;17:55.
7. Quince C, Ijaz UZ, Loman N, et al. Extensive modulation of the fecal metagenome in children with crohn's disease during exclusive enteral nutrition. *Am J Gastroenterol.* 2015;110:1718–29.
8. Reichardt N, Barclay AR, Weaver LT, et al. Use of stable isotopes to measure the metabolic activity of the human intestinal microbiota. *Appl Environ Microbiol.* 2011;77:8009–14.
9. Godon JJ, Zumstein E, Dabert P, et al. Molecular microbial diversity of an anaerobic digester as determined by small-subunit rDNA sequence analysis. *Appl Environ Microbiol.* 1997;63:2802–13.
10. Gerasimidis K, Bertz M, Hanske L, et al. Decline in presumptively protective gut bacterial species and metabolites are paradoxically associated with disease improvement in pediatric Crohn's disease during enteral nutrition. *Inflamm Bowel Dis.* 2014;20:861–71.
11. Caporaso JG, Lauber CL, Walters WA, et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci USA.* 2011;108(Suppl 1):4516–22.
12. Joshi NA, Fass JN. Sickle: A sliding-window, adaptive, quality-based trimming tool for fastq files. Version 1.21; 2011.
13. Masella AP, Bartram AK, Truszkowski JM, et al. PANDAseq: paired-end assembler for illumina sequences. *BMC Bioinformatics.* 2012;13:31.
14. Wang Q, Garrity GM, Tiedje JM, et al. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol.* 2007;73:5261–7.
15. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 2013;30:772–80.
16. Price MN, Dehal PS, Arkin AP. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS ONE.* 2010;5:e9490.
17. Oksanen J, Blanchet F G, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Wagner H. *vegan: Community Ecology Package*, R Package version 2.2-1, version 2.2-1 ed; 2015.
18. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE.* 2013;8:e61217.
19. Paradis E, Claude J, Strimmer K. APE: analyses of Phylogenetics and Evolution in R language. *Bioinformatics (Oxford, England).* 2004;20:289–90.
20. Schliep KP. phangorn: phylogenetic analysis in R. *Bioinformatics (Oxford, England).* 2011;27:592–3.
21. Cardoso P, Rigal F, Borges PAV, et al. A new frontier in biodiversity inventory: a proposal for estimators of phylogenetic and functional diversity. *Methods Ecol Evol.* 2014;5:452–61.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

